Factor-Timing Model

Ronald Hua, Dmitri Kantsyrev, and Edward Qian

Ronald Hua
is a chief investment officer at Goldman Sachs Asset Management in New York, NY.
ronald.hua@gs.com

Dmitri Kantsyrev
is a director of Equity at PanAgora Asset Management in Boston, MA.
dkantsyrev@panagora.com

Edward Qian
is the chief investment officer of Multi Asset at PanAgora Asset Management in Boston, MA.
eqian@panagora.com

Since the Great Recession of 2007 and the subsequent financial crisis of 2008, global financial markets have entered into an unchartered territory characterized by extreme macroeconomic conditions, elevated volatility, heightened correlations across multiple markets, and uncertain monetary and fiscal policy responses. In this environment, some of the traditional static quantitative equity strategies have struggled, in part due to the perverse behavior of their quantitative factors.

In a static quantitative model, the factor weightings are based on long-term risk-return statistics and show little change over a short time period. Therefore, while static models could ultimately perform well over the long term, they are vulnerable to changes in market conditions that may have an adverse impact on the model or factor performance in the short run.

As market volatility persists, static models could face performance difficulties. In this article, we suggest that investors can benefit from changes in market conditions by employing factor-timing strategies, transforming market volatility into an additional source of excess return. We propose a novel factor-timing approach that extends a static model framework and illustrates dynamic model-weight construction.

The major difference between a traditional static model and the dynamic factor-timing model is that the latter's factor weights are conditioned on a set of market variables and fluctuate with the values of conditioning variables through time. For example, the S&P 500 implied volatility index (VIX) can be a conditioning variable. When the VIX index is relatively high, one may take a contrarian view and buy stocks with a higher risk profile and short stocks with strong momentum. When the VIX index is relatively low, the reverse might be desirable. Whether or not this strategy is valid depends on the degree of the VIX's influence on expected returns and variances of momentum, value, or other quantitative factors.

The weights of a factor-timing model depend on both the conditional expected returns and the conditional return covariance, given the current conditioning variable readings. As conditioning variables vary over time, both conditional factor returns and the conditional covariance matrix change, creating a dynamic process of determining optimal model weights.

Starting with a given static model, there are two additional building blocks required for constructing a factor-timing model. The first is a set of optimal conditioning variables and the second is a process for assigning model weights to factors based on the observed values of conditioning variables in the optimal set.

In this article, the second step of determining model weights is similar to the static model framework discussed in Qian et al. [2004], in which the conditional expected returns and the conditional covariance are used in lieu of the unconditional ones.[1] Our main contribution in this article is to propose a framework for selecting conditioning variables in the first step. We also show how to track improvements in the resulting model's efficacy by deriving useful information ratios that demonstrate how a dynamic model adds value.

This article differs from the existing literature, as we aim to establish a unified methodology for constructing factor-timing models based on a conditioning information set. In contrast, the published literature is mainly concerned with employing conditioning variables to develop dynamic asset allocation strategies.[2]

OPTIMAL WEIGHTS FACTOR-TIMING MODEL WEIGHTS

To derive optimal dynamic factor-timing model weights, we extend Qian's et al. [2004] result by replacing the unconditional covariance matrix and the unconditional expected return vector with their conditional counterparts.

Assume R_{t+1} is a N-by-1 vector of random variables representing time-series returns of N factors, and V_t is a K-by-1 vector of random variables representing time-series readings of K conditioning variables. (For notational simplicity, we shall omit time subscripts for R and V from now on.) For example, while R contains returns to book-to-price and momentum factors, V includes VIX and other macroeconomic factors, such as the consumer confidence index. We further assume that both R and V follow a multivariate normal distribution, with the joint distribution of R and V expressed as

$$\begin{bmatrix} R \\ V \end{bmatrix} \sim N\left[\begin{pmatrix} \bar{R} \\ \bar{V} \end{pmatrix}, \begin{pmatrix} \Sigma_{RR} & \Sigma_{RV} \\ \Sigma_{VR} & \Sigma_{VV} \end{pmatrix}\right] \tag{1}$$

Then, by standard theory of conditional distribution, the conditional mean and conditional covariance of R, given a realization of V denoted by v, can be written as

$$\begin{aligned} R_{|v} &= \bar{R} + \Delta R \\ \Sigma_{|v} &= \Sigma_{RR} - \Sigma_{\Delta\Delta} \end{aligned} \tag{2}$$

where $\Delta R = \Sigma_{RV}\Sigma_{VV}^{-1}(v - \bar{V})$ is an adjustment to the return and $\Sigma_{\Delta\Delta} = \Sigma_{RV}\Sigma_{VV}^{-1}\Sigma_{VR}$ is an adjustment to the covariance. Given Equation (2), the conditional optimal model weights according to Qian et al. [2004], are $M_{|v}^{*} = \lambda\Sigma_{|v}^{-1}R_{|v} = \lambda\left(\Sigma_{RR} - \Sigma_{\Delta\Delta}\right)^{-1}\left(\bar{R} + \Delta R\right)$, in which λ is an arbitrary constant. Thus, the differences between static and dynamic model weights can be traced to two sources: the change in the covariance $\Sigma_{\Delta\Delta}$ and the change in the expected return ΔR that depends on realizations of conditioning variables relative to their means and is normally distributed with zero mean and covariance matrix $\Sigma_{\Delta\Delta}$.

CONDITIONING VARIABLE SELECTION

There is a large body of model selection literature that seeks to rank models in the candidate space relative to one another.[3] In this strand of literature, models are defined generally to represent their specifications, such as the variable selection, parameter estimation, and nesting structure. It is well recognized that the suitability of candidate models depends on a priori thinking. Given that financial markets are too complex to be replicated with a fully specified model, one can only hope to find a set of good approximating models that shed insights into markets' dynamics and inner workings.

Finding the set of candidate models with a priori knowledge is beyond the scope of this article. A successful exploration often depends on one's experience, knowledge, and creativity. Nonetheless, when developing a set of candidate variables, one should recognize the balancing act of keeping the set small by focusing on credible hypotheses and making it big enough not to omit a suitable a priori model. If a model does not make economic sense, don't include it. On the other hand, if one omits a potentially suitable variable, it cannot be rediscovered by a model selection algorithm. Make every effort to find the right balance.

Including more conditioning variables increases fitted-model precision and lowers in-sample residual variance. Yet, despite the improved data fit, adding variables may increase the forecast error's out-of-sample variance and reduce the fitted model's tractability. This article provides a selection mechanism based on Akaike's information criterion to evaluate various factor-timing specifications among a predetermined candidate set of variables. For instance, our framework allows us to answer the question, "How many conditioning variables should be included from the candidate set?"

AKAIKE'S INFORMATION CRITERION

Since the concept of entropy was first introduced in thermodynamics in the 19th century, it was well adopted across other disciplines, including information theory. In 1951, Kullback and Leibler conceptualized a measure of discrepancy between two models. Later, Akaike [1973] found a relationship between Kullback-Leibler's distance and Fisher's maximized log-likelihood function and proposed a methodology for selecting a parsimonious model. Akaike called it an *information criterion (AIC)*

$$AIC = -2\log(L) + 2\kappa \tag{3}$$

where L is the likelihood function for the estimated model and κ is the number of parameters. This measure places a premium on achieving a given fit with a smaller number of parameters per observation. It provides a simple way to select a model based on a data set. Specifically, one should pick a model that yields the smallest AIC value. The lower the AIC, the closer the model is to the unknown reality that generated the data.

To obtain AIC for a factor-timing model, assume T observations in the data sample with the distribution specified above. Then, as shown in the appendix

$$AIC = T \cdot \log\left[\left|\Sigma_{|\nu}\right|\right] + 2NK \tag{4}$$

That is, the AIC of a factor-timing model depends on the number of observations, T, the determinant of the conditional covariance matrix, $\left|\Sigma_{|\nu}\right|$, the number of quantitative factors, N, and conditioning variables, K.

When the number of observations T and the number of factors N are given, minimizing AIC means minimizing a linear combination of the number of conditional factors K and the logarithm of the determinant of the conditional covariance matrix. These two terms behave in a conflicted way: The higher the K, the better the fit and thus, the lower the determinant. Therefore, one cannot simply include many conditional variables, as a larger K lowers the determinant but raises the value of the second term in Equation (4).

A few interesting properties of AIC can be gleaned from of Equation (4). First, if we ignore correlations among residual returns and assume they are uncorrelated with each other, $\Sigma_{|\nu}$ is reduced to a diagonal matrix of conditional variances of factor returns and Equation (4) becomes

$$AIC \approx T \cdot \log\left[\prod_{n=1}^{N}\sigma_n^2\right] + 2NK = 2T\log\left(\prod_{n=1}^{N}\sigma_n\right) + 2NK \tag{4a}$$

As a logarithm is a monotonic function, the best model with the same number of conditioning variables K is the one with the lowest product of residual variances.[4]

Second, if we allow correlations among conditional factor returns and hold the number of conditioning variables K constant, the lower the determinant of the conditional covariance matrix, $|\Sigma_{|\nu}|$ and the better the model. Since $\Sigma_{|\nu}$ is a covariance matrix, it is semi-positive definite. At the extreme, the determinant is zero, AIC becomes infinitely negative, and the model is the absolute best.

There are two instances when the determinant can be zero. The first and somewhat trivial case is when one or more of the residual variances is zero. It means one can forecast factor returns with perfect accuracy, yielding a risk-free arbitrage. In the second case, $\Sigma_{|\nu}$ is linearly dependent with a rank less than N. One can again create a risk-free strategy by eliminating risk through offsetting factor returns.

Obtain one more interesting insight by transforming Equation (4) with eigenvalues. Assume that λ_n is the nth eigenvalue of the conditional covariance matrix, $\Sigma_{|\nu}$. Because the determinant of a matrix equals the product of its eigenvalues, i.e., $\left|\Sigma_{|\nu}\right| = \prod_{n=1}^{N}\lambda_n$, we can derive the following

$$AIC \approx T \cdot \sum_{n=1}^{N}\log(\lambda_n) + 2NK \tag{4b}$$

Equation (4b) shows that when AIC is used as a conditioning variables selection criterion, it favors variables that provide the largest *percentage* reduction of eigenvalues of the conditional covariance matrix, $\Sigma_{|\nu}$. It tends to select conditioning variables that can explain away a significant portion of common variations of factor returns, R.

Our discussion has been solely focused on how AIC selects the optimal factor-timing model based on the risk reduction. What about conditional factor returns, $R_{|v}$? They should also play a role in determining model efficacy. We will address this question in detail below, but for now note that risk reduction and conditional factor returns are the two sides of the same coin.

STEPWISE CONDITIONING VARIABLE SELECTION

Equipped with AIC in Equation (4), we devise a stepwise framework that selects conditioning variables one at a time. The proposed iterative procedure sequentially populates the optimal set of conditioning variables, S_i, until AIC reaches a trough and no longer decreases. Let i denote the current iteration of the process.

1. At the starting point, the *optimal* set of conditioning variables, S_0, is set to null and contains no variable. AIC_0 is equal to $T \cdot \ln\left[\left|\Sigma_{RR}\right|\right]$.
2. For each candidate variable k from the candidate set, compute the resulting model's AIC by including it into the S_i using Equation (4).
3. If a lower AIC value is obtained in step two than AIC_i, then the process continues to step four. Otherwise it stops, as the optimal set of conditioning variables has been found.
4. The variable k that resulted in the lowest AIC is added into S_i,

$$S_{i+1} = S_i \cup \{k\}$$

AIC_{i+1} is AIC of the model using S_{i+1} as the set of conditioning variables.
5. Continue to step two.

INFORMATION RATIO QUANTITIES

Since its introduction by Treynor and Black [1973], *Information Ratio (IR)*—the ratio of average excess return to the standard deviation of excess return—is commonly used in active portfolio management to evaluate a manager's forecasting efficacy. Therefore, it is important to provide a link between IR and AIC, the model-selection criteria introduced in the previous section.

Here, we derive four different IR quantities that can be used to track improvements in the model efficacy and show their dependence on changes from both the conditional return and the conditional covariance.

CONDITIONAL MODEL IR

First, we introduce the model information ratio at a point in time when we know the values of conditional variables. As we show in the appendix, *IR* in general is related to the average factor return and factor return covariance by

$$IR = (R'\Sigma^{-1}R)^{1/2} \tag{5}$$

Using the conditional factor return and conditional covariance, we obtain the conditional model IR as

$$Q = IR_{|v} = \sqrt{R'_{|v}\Sigma^{-1}_{|v}R_{|v}} \tag{6}$$

where the conditional return, $R_{|v} = \bar{R} + \Delta R$, and the conditional covariance, $\Sigma_{|v} = \Sigma_{RR} - \Sigma_{\Delta\Delta}$, are given in Equation (2). Since $\bar{R}$ and Σ_{RR} do not depend on realizations of conditioning variables, they are constants. Thus, the conditional model IR is a function of the adjustment to the conditional return, ΔR, and the decrease in the conditional covariance, $\Sigma_{\Delta\Delta}$.

As $\Sigma_{\Delta\Delta}$ depends on the optimal set of conditioning variables, the conditional model IR also depends on that set. In general, Q increases as the decrease in the conditional covariance $\Sigma_{\Delta\Delta}$ grows. More importantly, Q depends on the current readings of conditioning variables, v, as ΔR is equal to $\Sigma_{RV}\Sigma^{-1}_{VV}(v - \bar{V})$. As these readings change over time, Q varies accordingly and so behaves as a point-in-time measure.

EXPECTED MODEL IR

We define the *expected model IR* $E_{\Delta R}[Q]$ as the expected model IR through time for all possible adjustments to the conditional return ΔR. We derive the expected model IR by averaging the conditional IR, Q, over the distribution of ΔR.

Because $E_{\Delta R}[Q]$ does not have a closed-form solution, we first use the Taylor expansion to derive an approximation to the conditional IR. As a function of the conditional return, $R_{|v}$, Q can be expressed as

$$IR_{|\nu}(R_{|\nu}) = IR_{|\nu}\left[\bar{R}\right] + \left\{\frac{\partial\left(IR_{|\nu}\right)}{\partial R_{|\nu}}\left[\bar{R}\right]\right\}' \cdot \Delta R + \frac{1}{2}\Delta R' \cdot \left\{\frac{\partial^2\left(IR_{|\nu}\right)}{\partial (R_{|\nu})^2}\right\}\left[\bar{R}\right] \cdot \Delta R + O\left[\left\|\Delta R\right\|^3\right]$$
$$= \tau + \eta \cdot \Delta R + \frac{1}{2}\Delta R' \cdot \Theta \cdot \Delta R + O\left[\left\|\Delta R\right\|^3\right]$$

where

$$\tau = \sqrt{\bar{R}'\Sigma_{|\nu}^{-1}\bar{R}}, \eta = \tau^{-1}\left(\bar{R}'\Sigma_{|\nu}^{-1}\right), \Theta = \tau^{-3}\left(\bar{R}'\Sigma_{|\nu}^{-1}\bar{R}\cdot\Sigma_{|\nu}^{-1} - \Sigma_{|\nu}^{-1}\bar{R}\bar{R}'\Sigma_{|\nu}^{-1}\right) \tag{7}$$

Equation (7) is a quadratic approximation of the point in time measure, where η is the partial derivative of IR with respect to ΔR and Θ is the Hessian matrix of second-order partial derivatives. This approximation reveals how conditioning variables influence the conditional model information ratio. There are three sources of changes in the information ratio, and we discuss them in turn.

The constant term: risk reduction. The baseline model IR is $\sqrt{\bar{R}'\Sigma_{|\nu}^{-1}\bar{R}}$, the constant term τ of the Taylor expansion. Thus, with the expected return $\bar{R}$ being the same as in a static model, the baseline IR improvement is only due to the risk reduction. The conditional covariance matrix $\Sigma_{|\nu}$ is the residual of the unconditional one Σ_{RR} after removing the explained portion $\Sigma_{\Delta\Delta}$. In all likelihood, this IR is larger than the unconditional IR, $\sqrt{\bar{R}'\Sigma_{RR}^{-1}\bar{R}}$.

The linear term: interaction between $\bar{R}$ and ΔR. The linear term of the expansion $\eta\Delta R$ is $\tau^{-1}\cdot\bar{R}'\Sigma_{|\nu}^{-1}\Delta R$. Since τ is positive and does not depend on ΔR, the linear term increases or decreases the information ratio depending on how the average return $\bar{R}$ and the adjustment to the conditional return ΔR interact with one another.

The quadratic term: timing influence. The influence of the quadratic term $\Delta R'\Theta\Delta R$ on the information ratio depends on ΔR squared. Because Θ is semi-positive definite (as demonstrated in the appendix), the quadratic term is always non-negative and attains its smallest value—zero—when the adjustment to the conditional return ΔR is either zero or a constant multiple of the average return vector, $\bar{R}$. Therefore, this term tends to always improve the conditional IR, regardless of how the conditional variables behave. In a way, the factor-timing model adjusts itself to changes in the conditional factor return ΔR (due to the conditional variables), such as the factor exposures of the model are in line with ΔR.

Now we derive the expected model information ratio by integrating the approximation given by Equation (7) over the distribution of ΔR. As we show in the appendix,

$$H = E_{\Delta R}\left[Q\right] \approx \sqrt{\bar{R}'\Sigma_{|\nu}^{-1}\bar{R}} + \frac{1}{2}tr\left[\Theta\times\Sigma_{\Delta\Delta}\right] \tag{8}$$

In Equation (8), $\Sigma_{\Delta\Delta}$ is the covariance matrix of ΔR, and *tr* denotes the trace of a square matrix, or the sum of its diagonal elements. Notice that the linear term of the Taylor expansion drops out as $E_{\Delta R}[\Delta R]$ equals zero. This means that the linear term is only important to the point-in-time model IR, but not to the expected model IR.

The fact that the expected information ratio is only a function of the adjustment to the conditional covariance, $\Sigma_{\Delta\Delta}$, and does not depend on realizations of conditioning variables (that determine ΔR), validates the use of AIC as a selection criteria for factor-timing models, even though AIC only focuses on the model risk reduction.

As we discussed earlier, AIC gravitates toward conditioning variables that produce the lowest conditional covariance, $\Sigma_{|\nu} = \Sigma_{RR} - \Sigma_{\Delta\Delta}$. According to Equation (8), it increases the expected model IR, since both terms rise as the uncertainty in $\Sigma_{|\nu} = \Sigma_{RR} - \Sigma_{\Delta\Delta}$ lessens. Therefore, it is important for a modeler to focus on the global optimality (expected IR) rather than the local optimality (point-in-time IR).

EXPECTED MODEL IR SQUARED

Although there is no closed-form solution for the expected model IR, we can easily derive the expected value of information ratio squared. It is not a direct measure of model efficacy, but it does provide further insights on how a factor-timing model can be improved. Let us define quantity X as the expected model IR squared. We show in the appendix,

$$X = E_{\Delta R}\left[Q^2\right] = \mathrm{tr}\left[\Sigma_{|\nu}^{-1}(\bar{R}\bar{R}' + \Sigma_{\Delta\Delta})\right] \tag{9}$$

Equation (9), unlike Equation (8), is not an approximation. Once again it demonstrates that conditioning variables add value by both risk reduction and return enhancement. The former is represented by the $\Sigma_{|\nu}^{-1} = \left(\Sigma_{RR} - \Sigma_{\Delta\Delta}\right)^{-1}$ term; the latter is represented by $(\bar{R}\bar{R}' + \Sigma_{\Delta\Delta})$. As a result of conditioning, the former term gets smaller and the latter increases by the amount of the decrease in the conditional covariance $\Sigma_{\Delta\Delta}$. Thus, an alternative information ratio quantity, IR squared, further supports the application of AIC for building a factor-timing model.

MULTI-PERIOD IR

In the portfolio-management practice, investment managers and consultants frequently use realized multi-period model-information ratios to evaluate a strategy's added value. We must emphasize that the multi-period IR is different from the expected IR. Empirically, we can compute the multi-period IR, denoted by P, by performing the following steps.

First, calculate model returns in each period. Then compute the multi-period IR as the ratio of the average return to the standard deviation of returns. Statistically, this is the ratio of two expectations.

In contrast, the average model IR – $E(Q)$ simply estimates the average value of single-period information ratios. In statistical terms, it is the expectation of the ratio of return to standard deviation. The difference is subtle, but the ratio of the two expectations is not equal to the expectation of the ratio, i.e., $E(x)/E(y) \neq E(x/y)$.

Confusion in the investment industry regarding the fundamental law of active management has long been rooted in the mistake of treating the average of single-period information ratios as a multi-period IR. Exhibit 1 presents a graphical illustration of the difference between the two quantities.

To compute P, we first need to derive the functional form of model returns in each period - a_t,

$$a_t = r_t' \times M_t = \lambda \cdot r_t' \times \Sigma_{|\nu}^{-1} R_{|\nu,t} = \lambda \cdot (R_{|\nu,t} + \varepsilon_t)' \Sigma_{|\nu}^{-1} R_{|\nu,t} \tag{10}$$

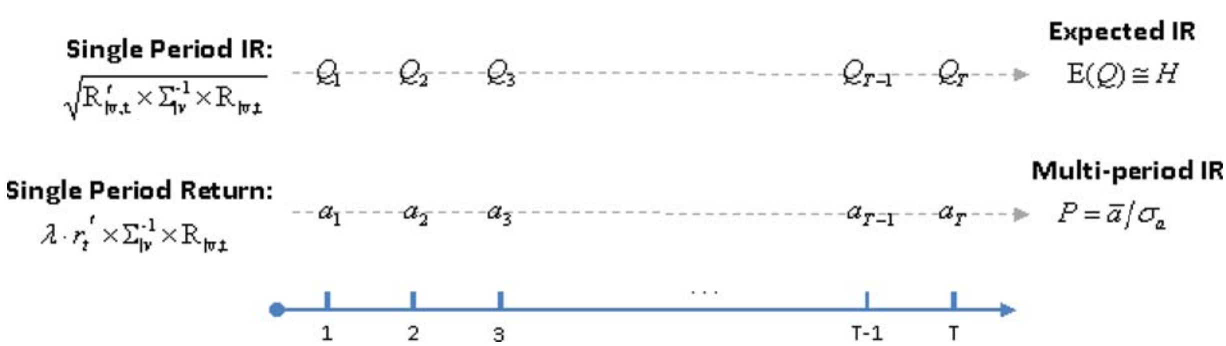
where r_t' and M_t represent realized factor returns and conditional model weights defined in Equation (2) respectively, and $\Sigma_{|\nu}$, $R_{|\nu,t}$, and ε_t denote conditional covariance, conditional returns, and residual returns at time t respectively. Then, we can find the multi-period IR as

$$P = \frac{\bar{a}}{\sigma_a} = \frac{E(a)}{\sqrt{E(a^2) - E(a)^2}} = \frac{X}{\sqrt{E(a^2) - X^2}} \tag{11}$$

where X is the expected model IR squared given by Equation (9). Note that $E(a)$ equals λX as the expected value of $R_{|\nu}\varepsilon'$ is a zero matrix.

Unfortunately, the expectation of return squared eludes an analytic solution. To estimate the exact value of multi-period IR, one must compute it numerically. Although the multi-period IR is the correct industry standard for calculating the information ratio quantity and evaluating strategies' added values, the expected

EXHIBIT 1
Expected IR vs. Multi-Period IR



model IR has the advantage of being tractable. Its derivation elucidates how factor-timing models add value.

AN EMPIRICAL EXAMPLE

To illustrate the proposed framework, we provide an empirical example of how to select conditioning variables using Akaike's information criterion, and how to track improvements in the efficacy of resulting models.

Our sample period spans from January 1994 to May 2009, with a forecast horizon of one month forward. There are 185 independent monthly periods in the sample. The security universe consists of U.S. companies in the Citigroup Primary BMI index. On average, it includes 410 companies with the market's largest capitalization each month. Fundamental and pricing data come from Worldscope and IDC databases.

DESCRIPTION OF FACTORS AND CONDITIONING VARIABLES

We choose three factors: earnings yield (E2P), return on equity (ROE), and six-month price momentum (PM6).[5] Two reasons prompt us to use these factors in our illustration.

First, among the diverse set of factors commonly used to predict future returns, these three tend to have higher strategy risk (higher standard deviations of factor returns through time) and lower strategy return (lower average factor returns through time). A relatively high strategy risk provides an opportunity to add value through factor timing. In contrast, factors with relatively low strategy risks might be more robust, and thus more appropriate for static models.

Secondly, each of these three factors represents one of the three factor categories frequently used in quantitative equity models. These categories are value (buying cheap stocks), quality (buying good companies), and momentum (riding market sentiment).

The candidate set of conditioning variables consists of the S&P 500 implied volatility index (VIX), the trailing twelve-month S&P 500 index return (SPX), the Fed funds rate (FED), the U.S. consumer confidence index (CC), the calendar month (CAL), the debt-to-market capitalization ratio spread (D2M), and the spread of book-to-price ratio (B2P).[6] All of these conditioning variables have the potential to time various factors, as has been documented in many studies.

For example, Copeland and Copeland [1999] found that changes in the volatility index differentiate future daily returns for size and value factors. They show that on days that follow increases in the VIX, portfolios of large-capitalization stocks and value-based portfolios outperform their small and growth counterparts. Cooper et al. [2004] showed that momentum profits depend on the state of the market. The mean monthly momentum profit following positive market returns is positive; it is negative following negative market returns.

A number of papers have also examined the association of monetary policy-related variables with market or style returns. For example, Conover et al. [2005] find that changes in the discount rate and the associated monetary policy environment—restrictive or expansive—have a strong relationship with security returns, and that small-capitalization companies are more sensitive to changes in monetary conditions. Similarly, Chordia and Shivakumar [2002] argued that returns to momentum strategies are related to the business cycle: They are positive only during expansionary periods. Finally, Asness et al. [2000] used the value spread and the earnings-growth spread as explanatory variables to show that value-growth style returns may be predictable.

To avoid forward-looking bias, each month current conditioning variables' readings are ranked in a uniform distribution relative to their prior sixty observations. Then the uniform distribution is scaled to range from 0 to 100, with 100 being the highest data point and 0 being the lowest data point in the preceding five years. (For CAL we use numbers one through twelve to represent January through December). We use these ranks to time factor returns.

Exhibit 2 presents summary statistics computed over the considered time period. For factors, means are averages of rank correlations between factor values and one-month forward total returns (average rank IC). The covariance matrix presents variances and covariances of these rank correlations among themselves in the left box, as well as their covariances with conditioning variables in the right box. For conditioning variables, means are averages of conditioning variables ranks. The covariance matrix shows variances and covariances of these ranks in the right box. Note that because trailing comparison windows move forward with each new observa-

EXHIBIT 2
Summary Statistics

			Covariance Matrix									
		Mean	**ROE**	**E2P**	**PM6**	**B2P**	**D2M**	**FED**	**CC**	**SPX**	**VIX**	**CAL**
Factors	ROE	0.02	0.02	0.01	0.01	–0.17	–0.24	0.24	0.25	0.12	0.32	–0.01
	E2P	0.03	0.01	0.03	0.00	–0.70	–0.92	0.21	0.02	0.65	–0.02	–0.02
	PM6	0.02	0.01	0.00	0.04	–0.47	–0.61	0.47	0.84	–0.41	–0.08	0.01
Conditioning Variables	B2P	57.3	–0.17	–0.70	–0.47	782.7	639.0	–45.5	–79.2	250.8	279.8	–6.2
	D2M	55.0	–0.24	–0.92	–0.61	639.0	773.5	–88.0	–32.6	191.3	307.4	2.6
	FED	38.3	0.24	0.21	0.47	–45.5	–88.0	664.3	508.1	–335.9	–17.6	6.7
	CC	51.4	0.25	0.02	0.84	–79.2	–32.6	508.1	1025.2	–622.2	340.3	–4.8
	SPX	57.6	0.12	0.65	–0.41	250.8	191.3	–335.9	–622.2	878.2	–41.8	0.8
	VIX	55.4	0.32	–0.02	–0.08	279.8	307.4	–17.6	340.3	–41.8	1137.3	2.2
	CAL	6.5	–0.01	–0.02	0.01	–6.2	2.6	6.7	–4.8	0.8	2.2	11.8

EXHIBIT 3
Conditioning Variables in the First Iteration

			IR Quantiities				
Step	**Conditioning Variable**	**AIC**	**Q**	**H**	**X**	**E(Q)**	**P**
1	B2P	–570.3	0.614	0.262	0.092	0.272	0.288
1	**D2M**	**–576.4**	**0.460**	**0.295**	**0.128**	**0.317**	**0.288**
1	FED	–564.2	0.106	0.219	0.057	0.221	0.222
1	CC	–565.1	0.177	0.238	0.062	0.238	0.268
1	SPX	–566.0	0.263	0.250	0.067	0.251	0.257
1	VIX	–564.7	0.249	0.244	0.059	0.242	0.239
1	CAL	–562.6	0.217	0.216	0.048	0.216	0.219

tion, means and medians of conditioning variables ranks through time are unlikely to be exactly 50.

CONDITIONING VARIABLE SELECTION

Here we implement the iterative procedure described above that selects conditioning variables from the candidate set into the optimal set. In the first iteration, each conditioning variable is tested individually. Exhibit 3 presents AIC values and information ratios obtained by each model. Note that because Q is a point-in-time measure and depends on current readings of conditioning variables, it is shown for the last observation dated May 2009.

In addition to the four information ratios discussed above, we numerically compute the *average conditional model information ratio*—$E(Q)$, to gauge whether the expected model information ratio H obtained using the Taylor expansion provides a reasonable approximation. Indeed, Exhibit 3 reveals that H is a relatively good proxy for $E[Q]$: their rank orders are identical. In contrast, rank orders between $E[Q]$ and the multi-period IR P are different, with only four out of seven the same. Their ranked correlation is 0.89.

Because D2M has the lowest AIC value in Exhibit 3, −576.4, we choose it as the first conditioning variable. Then the procedure continues to iterate and more conditioning variables are selected into the optimal set. In each step, a conditioning variable is selected with the lowest value of AIC. Exhibit 4 shows that AIC decreases as new conditioning variables are added to the optimal set.

After the third step, iterations stop, as AIC is equal to −589.5 and is no longer decreasing. As shown in Exhibit 5, the remaining variables—B2P, FED, VIX, and CAL—produce AIC values that are greater than the value obtained in the third step. Even though information ratio quantities marginally improve if we include these variables, the increased complexity of resulting models no longer warrants those improvements. Also note that, even though B2P is the second-best condi-

EXHIBIT 4

Conditioning Variables Selected in Each Iteration

Step	Conditioning Variable	AIC	IR Quantiities Q	H	X	E(Q)	P
0		–564.2	0.212	0.212	0.045	0.212	0.212
1	D2M	–576.4	0.460	0.295	0.128	0.317	0.288
2	SPX	–581.7	0.598	0.348	0.171	0.357	0.309
3	CC	–589.5	0.354	0.366	0.233	0.401	0.285

EXHIBIT 5

Conditioning Variables in the Last Iteration

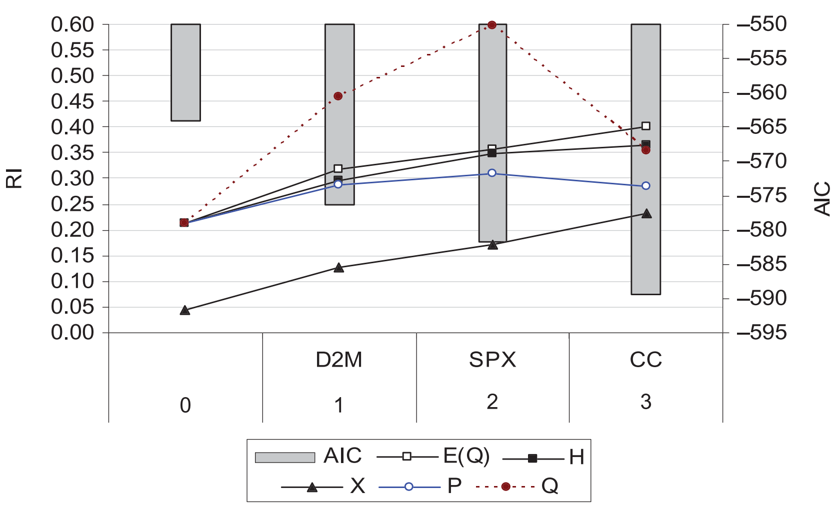
Step	Conditioning Variable	AIC	IR Quantiities Q	H	X	E(Q)	P
4	B2P	–587.6	0.403	0.367	0.234	0.402	0.287
4	FED	–588.0	0.326	0.368	0.236	0.403	0.279
4	VIX	–589.4	0.422	0.387	0.243	0.414	0.294
4	CAL	–587.9	0.357	0.371	0.235	0.404	0.291

tioning variable in the first iteration, it is not picked in this exercise. The reason: B2P has a high correlation with D2M (it is equal to $0.82 = 639/\sqrt{782.7 \cdot 773.5}$).

Although AIC solely focuses on risk reduction, it in effect selects variables that result in the highest-expected model information ratio. As shown in Exhibit 3, D2M—the conditioning variable selected in the first step—results in the highest measure of expected model IR. As AIC decreases with each additional iteration, the information ratio quantities H, X, and $E(Q)$ monotonically increase as illustrated in Exhibit 4. (The conditional information ratio Q fluctuates, as it is a point-in-time measure and depends on current readings of conditioning variables at the last point of our data). This pattern is also clearly observable in Exhibit 6. On the other hand, successive improvements in quantity P are smaller than those in $E[Q]$, and P decreases in the third iteration, rather than increasing.

EXHIBIT 6

AIC and Information Ratio Quantities



Two additional points can be gleaned from Exhibit 6. First, the expected model information ratio, H, is a good proxy for the expected model information ratio computed numerically, $E(Q)$, as the two lines move close to each other. Given the computational intensity required to calculate $E(Q)$, H seems to be an efficient approximation to track expected model IR changes.

Second, although the average of IR squared does not equal the average of IR raised to the power of two, $E(IR^2) \neq E(IR)^2$, the expected IR squared, X, appears to be a good measure to track the expected model IR changes. The correlation between X

and $E(Q)$ is 0.997, whereas the correlation between the approximation H and $E(Q)$ is 0.985.

SOURCES OF IR IMPROVEMENT

We have discussed the three sources of improvement in the conditional information ratio due to the introduction of conditioning variables. The example here lets us demonstrate the magnitude of those improvements. Consider the optimal model obtained in the third iteration, as shown in Exhibit 4.

The Constant Term: Risk Reduction

The introduction of conditioning variables reduces the covariance matrix from Σ_{RR} to $\Sigma_{|v}$, and thus increases both the conditional point-in-time and the expected model information ratios from $(\bar{R}'\Sigma_{RR}^{-1}\bar{R})^{1/2}$ to $(\bar{R}'\Sigma_{|v}^{-1}\bar{R})^{1/2}$. In our example, the risk reduction component improves the information ratio quantities by 5.7% (H increases from 0.212 to 0.224). It is a rather small improvement.

The linear term: interaction between $\bar{R}$ and ΔR. The linear term increases or decreases the conditional information ratio, depending on the interaction between the average return $\bar{R}$ and the adjustment to the conditional return ΔR at different points in time. However, since the average value of this term is zero, it does not change the expected model IR. Panel A of Exhibit 7 presents the distribution of the linear term. Indeed, while its contribution to the conditional IR varies from −0.653 to 1.181, depending on the point in time, its mean is zero.

The quadratic term: timing influence. This term is non-negative and, according to Equation (8), its expected value is $\frac{1}{2}\cdot tr\left[\Theta\times\Sigma_{\Delta\Delta}\right]$. Panel B of Exhibit 7 shows the distribution of the quadratic term. Its contribution to the expected model IR is always positive, with an average value of 0.142. That is a 67% improvement from the static model.

To summarize, the increased model efficacy primarily comes from the quadratic term: ΔR squared. It is not a surprise that, as the adjustment to the conditional return ΔR varies, model weights follow and change their factor exposures, too. The dynamic weighting mechanism can potentially add significant value when ΔR correctly forecasts factor returns.

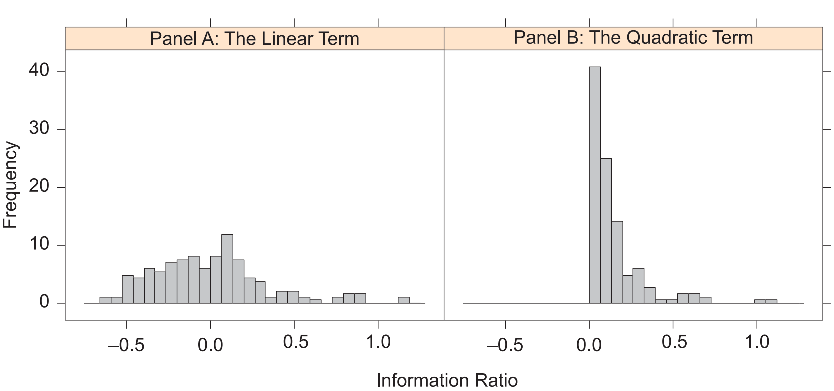
CONCLUSION

A large body of literature documents conditioning variables that can be used to time factor returns. Besides that, factor timing is a topic of constant interest among many professionals. However, few attempts have been made to introduce a unified methodology for constructing an optimal factor-timing model from a set of factors and conditioning variables. We propose a novel framework by extending results from Qian et al. [2004] around building an optimal static model, and by incorporating the concept of Akaike's [1973] information criterion. Furthermore, we derive useful information ratio quantities to track improvements in model efficacy; we also quantify sources of added value.

We believe that research in the area of dynamic model weighting is still in its infancy, and that our results can be extended in multiple directions. First, one of the most onerous assumptions in our derivations is that both factor returns and conditioning variables follow *multi-variate Gaussian* distribution. In practice this assumption is often violated. Perhaps a non-parametric framework can be devised as the counterpart to our parametric one.

Second, even though AIC seeks to minimize the chance of model over-fitting, our information ratio quantities are still in-sample measures and thus unrealistically optimistic. Using the concept of cross-validation or some form of blind-forward techniques may increase the like-

EXHIBIT 7
Histograms of IR Contributions



lihood of getting realistic estimations. This extension may be indispensible to the betterment of ex post model performance.

APPENDIX

In this appendix we provide technical details of several equations in the main text.

To derive Equation (4), let ε_t be a vector representing the tth observation of residuals of random variables R after conditioning on V. In addition, for expositional simplicity and analytic tractability, let us assume that ε_t is identically distributed and serially independen, for t = 1, 2, …, T. In general, the last assumption can be relaxed, but then one needs to use a partial maximum likelihood estimator and the estimation becomes computationally intensive. With our restrictive assumptions, the likelihood for the estimated model is

$$L(v) = \frac{1}{(2\pi)^{NT/2}\left|\Sigma_{|v}\right|^{T/2}} \cdot \exp\left[-\frac{1}{2}\cdot\sum_{t=1}^{T}\varepsilon_t'\Sigma_{|v}^{-1}\varepsilon_t\right] \quad \text{(A-1)}$$

The sum in the exponent can be simplified by approximating it as an expectation.

$$\sum_{t=1}^{T}\varepsilon_t'\Sigma_{|v}^{-1}\varepsilon_t \cong T\cdot E\left[\varepsilon'\Sigma_{|v}^{-1}\varepsilon\right] = TN \quad \text{(A-2)}$$

Therefore,

$$L(v) = \frac{1}{(2\pi)^{NT/2}\left|\Sigma_{|v}\right|^{T/2}} \cdot \exp\left(-\frac{NT}{2}\right) = (2\pi e)^{-NT/2}\left|\Sigma_{|v}\right|^{-T/2} \quad \text{(A-3)}$$

Because the number of estimated residual covariances is $N(N+1)/2$ and the number of estimated coefficients and intercepts is $N(K+1)$, we can derive AIC for a factor-timing model using Equation 3 as

$$\begin{aligned} AIC = NT\cdot\left[\log(2\pi)+1\right] + T\cdot\log\left[\left|\Sigma_{|v}\right|\right] \\ + N(N+1) + 2N(K+1) \end{aligned} \quad \text{(A-4)}$$

After removing constants $NT\cdot[\log(2\pi) + 1]$ and $N(N + 1) + 2N$ that do not change the rank order of AIC, we get Equation (4), i.e., $AIC = T\cdot\log[|\Sigma_{|v}|] + 2NK$.

To derive Equation (5), note that the optimal portfolio weights are proportional to the product of the inverse conditional covariance matrix and the expected conditional return (Qian et al. [2004]). Then, the model return R_m and variance σ_m^2 are

$$\begin{aligned} R_m &= M^{*\prime}R = \lambda_m\cdot R'\Sigma^{-1}R \\ \sigma_m^2 &= M^{*\prime}\Sigma M^* = \lambda_m^2\cdot R'\Sigma^{-1}\Sigma\Sigma^{-1}R = \lambda_m^2\cdot R'\Sigma^{-1}R \end{aligned} \quad \text{(A-5)}$$

Now, we can obtain the model information ratio as

$$IR_m = \frac{R_m}{\sigma_m} = \frac{\lambda_m\cdot R'\Sigma^{-1}R}{\lambda_m\cdot(R'\Sigma^{-1}R)^{1/2}} = \sqrt{R'\Sigma^{-1}R} \quad \text{(A-6)}$$

In Equation (7), Θ is semi-positive definite. To prove that, denote $A = (r'\Sigma r)\Sigma - \Sigma rr'\Sigma$ Then for any vector x we have $x'Ax = (r'\Sigma r)(x'\Sigma x) - (x'\Sigma r)$. Since Σ is positive definite, according to the Schwarz inequality - $(r'\Sigma r)(x'\Sigma x) \geq (x'\Sigma r)^2$, and the two sides are equal only when $x = r$. then $x'Ax \geq 0$ and A is semi-positive definite.

To derive Equation (8), we integrate Q over the conditional return $R_{|v}$. Since $\bar{R}$ is a constant, integrating over ΔR produces the same result,

$$H \approx E_{\Delta R}\left[\tau + \tau^{-1}\cdot\bar{R}'\Sigma_{|v}^{-1}\Delta R + \tfrac{1}{2}\Delta R'\Theta\Delta R\right] \quad \text{(A-7)}$$

The expectation of the linear term is zero and the expectation of the quadratic term gives rise to

$$\begin{aligned} H &\approx E_{\Delta R}\left[\tau + \tau^{-1}\cdot\bar{R}'\Sigma_{|v}^{-1}\Delta R + \tfrac{1}{2}\Delta R'\Theta\Delta R\right] \\ &= \tau + \frac{1}{2}tr\left\{\Theta\times E_{\Delta R}\left[\Delta R\Delta R'\right]\right\} = \tau + \frac{1}{2}\cdot tr\left\{\Theta\times\Sigma_{\Delta\Delta}\right\} \end{aligned} \quad \text{(A-8)}$$

For Equation (9), the average of IR squared can be obtained similarly,

$$\begin{aligned} X &= E_{\Delta R}\left[Q^2\right] = E_{\Delta R}\left[\left(\bar{R}+\Delta R\right)'\Sigma_{|v}^{-1}\left(\bar{R}+\Delta R\right)\right] \\ &= E_{\Delta R}\left\{\mathrm{tr}\left[\left(\bar{R}+\Delta R\right)'\Sigma_{|v}^{-1}\left(\bar{R}+\Delta R\right)\right]\right\} \\ &= \mathrm{tr}\left\{\Sigma_{|v}^{-1}E_{\Delta R}\left[\left(\bar{R}+\Delta R\right)\left(\bar{R}+\Delta R\right)'\right]\right\} \\ &= \mathrm{tr}\left\{\Sigma_{|v}^{-1}\left(\bar{R}\bar{R}'+\Sigma_{\Delta\Delta}\right)\right\} \end{aligned} \quad \text{(A-9)}$$

ENDNOTES

[1]As illustrated by Qian et al. [2004], we build a multi-factor model by employing traditional portfolio theory with the objective of maximizing the expected model-information ratio. Consequently, factor weights in our model are proportional to the expected factor returns and inversely proportional to the factor-return covariance. That is, a factor in the model commands a higher weight if it has higher return, lower volatility, or lower correlations with other factors.

[2]For example, Solnik [1993] proposed dynamic international allocation strategies that are based on a conditioning information set. Similarly, Harvey [1994] focused on conditional asset-allocation strategies in the emerging markets. In the most recent work, Sharpe [2010] advocated an asset allocation policy that adapts to market movements by taking into account changes in the market values of major asset classes.

[3]Some of the more popular and well-known models include F tests, stepwise, backward- and forward-selection procedures, bootstrap and cross-validation, Bayes factors, AIC, BIC, Mallows' C_p, and so on. Also see Burnham and Anderson [2002] for a detailed discussion on model selection.

[4]Equation (4a) demonstrates that AIC is a function of the product of residual variances. A geometric interpretation of $\prod_{n=1}^{N}\sigma_n$ is an area covered by N edges whose lengths are denoted as $(\sigma_1, \ldots, \sigma_N)$, standard deviations of the residual covariance matrix. The smaller the area, the better the AIC.

In the case of adding an additional conditioning variable, the magnitude of AIC improvement is directly linked to the reduction of the area covered by the standard deviations of the residual covariance matrix. Hence, it is the *percentage change* of the standard deviation that matters, not the magnitude of the standard deviation change. For example, suppose there are two factors: one with a residual return standard deviation of 100 and the other with a standard deviation of 1. Decreasing the standard deviation of the first factor from 100 to 50 will yield the same AIC improvement as decreasing the second factor standard deviation from 1 to 0.5.

[5]Each month, E2P is calculated as the twelve-month net income before extraordinary items, divided by the market capitalization. ROE is equal to the twelve-month net income before extraordinary items divided by the common equity twelve months ago, and PM6 is calculated as the total return that includes both change to the market capitalization and dividends payments between the last month and six months ago. Each month we calculate rank correlations between factors and one-month forward returns.

[6]At the end of each month, VIX, SPX, FED, and CAL are directly observable, CC is the most recent value known, and spreads of debt-to-market capitalization ratio and book-to-price ratio are calculated as the distance between median values of the first and tenth percentile of these ratios. Note that this is only an example. The selection of conditioning variables is nowhere near exhaustive.

REFERENCES

Akaike, H. "Information Theory as an Extension of the Maximum Likelihood Principle, in Petrov and Csake (eds.)." Second International Symposium on Information Theory, Academiai Kiado, Budapest, (1973).

Asness, C., J. Friedman, R. Krail, and J. Liew. "Style Timing: Value Versus Growth." *The Journal of Portfolio Management*, 26 (Spring 2000), pp. 50-60.

Burnham, K., and D. Anderson. "Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach." Second edition, New York: Springer-Verlag (2002).

Chordia, T., and L. Shivakumar. "Momentum, Business Cycle, and Time-varying Expected Returns." *Journal of Finance*, Vol. 57, No. 2 (2002), pp. 985-1019.

Conover, M., G. Jensen, R. Johnson, and J. Mercer. "Is Fed Policy Still Relevant for Investors?" *Financial Analysts Journal*, Vol. 61, No. 1 (2005), pp. 70-79.

Cooper, M., and R. Gutierrez, Jr. "Market States and Momentum." *Journal of Finance*, Vol. 59, No. 3 (2004), pp. 1345-1365.

Copeland M., and T. Copeland. "Market Timing: Style and Size Rotation Using the VIX." *Financial Analysts Journal*, Vol. 55, No. 2, (1999), pp. 73-81.

Harvey, C. "Portfolio Enhancement Using Emerging Markets and Conditioning Information, in Claessens and Gooptu (eds.)." Portfolio Investment in Developing Countries (1994).

Qian, E., E. Sorensen, R. Hua, and R. Schoen. "Multiple Alpha Sources and Active Management." *The Journal of Portfolio Management*, 30 (Winter 2004), pp. 39-45.

Qian, E., R. Hua, and E. Sorensen. "Quantitative Equity Portfolio Management: Modern Techniques and Applications." New York, NY: Chapman & Hall (2007), p. 195.

Sharpe, W. "Adaptive Asset Allocation Policies." *Financial Analysts Journal*, Vol. 66, No. 3 (2010), pp. 45-59.

Solnik, B. "The Performance of International Asset Allocation Strategies Using Conditioning Information." *Journal of Empirical Finance*, 1 (1993), pp. 33-55.

Treynor, J.L., and F. Black. "How to Use Security Analysis to Improve Portfolio Selection." *Journal of Business*, 46 (1973), pp. 66-86.